Jacopo Minniti

jacopominniti.com | jacopo.minniti004@gmail.com | Google Scholar | GitHub | LinkedIn

Research Overview

My research aims to develop a predictive theory of deep learning by combining mathematical analysis with empirical investigation. I study how neural networks acquire reusable computations during training, how composition and overparameterization shape learning dynamics, and how latent representations evolve during multi-step reasoning.

Education

Minerva University Sep 2023 – May 2027

Bachelor of Science in Computational Sciences

- Concentrations: AI & Mathematics. **GPA 4.00 (major); 3.94 (overall).**
- Relevant Coursework: Real Analysis, Group Theory, Linear Algebra, Optimization, Differential Equations, Probability & Statistics, Bayesian Statistics, Machine Learning.

Publications & Preprints

Neural Dynamics as the Composition of Quantized Units

Minniti J., Kulanthaivelu A., Sproat R. (2026).

Submitted to ICLR 2027.

- Developed a theory of learning as the acquisition of discrete computations, predicting their order and recovering scaling-law exponents within 1.4–5.7% relative error on controlled tasks.
- Recovered interpretable computations from Transformer training trajectories; quanta-guided training improved number-naming exact-sequence accuracy from 84.9% to 94.7%, and from 50.5% to 89.9% on short-tail examples.

Learned Relay Representations for Forward-Thinking Discrete Diffusion Models

Rozonoyer B.*, **Minniti J.***, Patel D.*, Band N., Bose A. J., Rudner T. G. J., McCallum A. (2026).

*ICML 2026 SPIGM Workshop; submitted to ICLR 2027. *Equal contribution.*

- Introduced continuous relay representations trained with backpropagation through time to preserve and reuse latent computations across diffusion denoising steps.
- Improved HumanEval accuracy from 38.4% to 42.1% on Fast-dLLM v2 (1.5B), while using 32% fewer denoising forward passes than standard supervised fine-tuning.

Challenges in Inference-Time Scaling with Uncertainty-Aware Tree Search

Minniti J., Band N., Rudner T. G. J. (2026).

ICLR 2026 Workshop on Scaling Post-Training for LLMs (SPOT).

- Developed uncertainty-aware tree search, dynamically allocating additional reasoning branches to states where process uncertainty estimates indicate greater ambiguity.
- Showed that uncertainty estimates correlated with rollout variance on held-out data ($r = 0.62$) became unreliable under search, with correlation dropping by 95% ($r = 0.03$).

Effects of Age, Semantic Relatedness, and Vocabulary Knowledge on New Word Meanings

Chen P.-H., Hulme R. C., **Minniti J.**, Lee C.-L., Rodd J. M. (2026).

Presented at CLDC 12; submitted to the Journal of Memory and Language.

- Investigated acquisition of new word meanings in 300+ participants, finding that semantic relatedness improved recall ($p < .001$) and vocabulary knowledge predicted stronger form–meaning associations.

Research Experience

Member of Technical Staff, *Sakana AI, Tokyo, Japan* May – Sep 2026

- Studied the emergence of reusable computations in neural networks through theoretical models, controlled experiments, and mechanistic analysis of Transformer training.

Research Assistant, *Tim G. J. Rudner Group, University of Toronto*, *Remote* Sep 2025 – May 2026

- Developed a framework to identify discrete computations learned by Transformers, predict their order of acquisition, and leverage this structure to improve generalization in low-data regimes.

AI Safety Researcher, *XLab, University of Chicago*, *Chicago, USA* Jun – Aug 2025

- Trained probes on language-model hidden states to predict step-level reasoning uncertainty, achieving up to 0.89 AUROC on MATH500.

Undergraduate Researcher, *Brain and Language Processing Lab, NTU, Taiwan* Sep – Dec 2024

- Analyzed EEG/ERP recordings from 50 participants to study linguistic expectation during sentence comprehension, examining N400/LPP responses and their relationship to GPT-2 surprisal estimates.

Teaching Experience

- Teaching Assistant, **Introduction to Optimization Theory**, Minerva University (Fall 2026).
- Teaching Assistant, **Real Analysis**, Minerva University (Spring 2026).
- Teaching Assistant, **Fundamentals of Programming in Python**, Minerva University (Fall 2024).

Talks

- Minerva University, **Fault Tolerant Distributed Training in Torch** (2026).
- CLDC 12, **Semantic Relatedness and Vocabulary Knowledge in New Word Meanings** (2026).
- University of Chicago: **The (Not So) Bitter Lesson** (2025).

Technical Projects

Neural Option 2026

- Developed a model of how learning frequent features can weaken alternative responses.
- Predicted which learned feature distinctions would weaken during training, achieving 0.94 AUC on a parity task.

Convergent Minds 2026

- Built a PyTorch framework for aligning brain activity with language models, including leakage-safe preprocessing, neural decoding architectures, and reusable evaluation pipelines.

Perceptual Gate for Continuous Thought Machines 2025

- Added a learnable gate between sensory input and recurrent latent computation; improved CIFAR-10 accuracy by 4 percentage points and adaptive computation by 12.

Honors & Awards

- **Top 1% by GPA** of Minerva University's Class of 2027
- **School on Analytical Connectionism** (2026): Seminar on neural networks theory (only undergraduate participant in a graduate-level program).
- **Mistral AI Hackathon, San Francisco** (2024): Selected "top projects" for interactive video learning assistant
- **Cal Hacks, UC Berkeley** (2023): Amazfit and MindsDB prizes for a wearable-based emotional wellness tool.
- **NASA Space Apps, Mountain View** (2023): First place winner for an AI-based wildfire evacuation system.

Technical Skills

- **Programming Languages:** Python, Java, JavaScript, Dart, MATLAB, SQL.
- **ML Areas & Methods:** Mechanistic interpretability, reinforcement learning and post-training, stochastic processes, dynamical systems, computational neuroscience.
- **Research Tools:** PyTorch, JAX, NumPy, SciPy, Matplotlib, EEGLAB.
- **Systems & Infrastructure:** CUDA, Slurm, Docker, Git, AWS (EC2, S3, SageMaker).